

Evolutionary relationship of alfalfa mosaic virus with cucumber mosaic virus and brome mosaic virus

H. S. SAVITHRI and M. R. N. MURTHY*

Department of Biochemistry and *Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012

MS received 5 August 1983

Abstract. The amino acid sequences of the non-structural protein (molecular weight 35,000; 3a protein) from three plant viruses – cucumber mosaic, brome mosaic and alfalfa mosaic have been systematically compared using the partial genomic sequences for these three viruses already available. The 3a protein of cucumber mosaic virus has an amino acid sequence homology of 33.7% with the corresponding protein of brome mosaic virus. A similar protein from alfalfa mosaic virus has a homology of 18.2% and 14.2% with the protein from brome mosaic virus and cucumber mosaic virus, respectively. These results suggest that the three plant viruses are evolutionarily related, although, the evolutionary distance between alfalfa mosaic virus and cucumber mosaic virus or brome mosaic virus is much larger than the corresponding distance between the latter two viruses.

Keywords. Brome mosaic virus; cucumber mosaic virus; alfalfa mosaic virus; sequence homology; evolutionary relationship.

Introduction

Plant viruses are grouped into different classes based on a detailed examination of their host and vector specificity, serological relationships, particle morphology, nucleic acid content and nature, and genomic organization. Alfalfa mosaic virus (AMV), cucumber mosaic virus (CMV) and brome mosaic virus (BMV) are classified into different families of plant viruses, (Matthews, 1979) although they have a similar genomic organization. AMV (Jaspars, 1974), CMV (Kaper and Waterworth, 1981) and BMV (Lane, 1981) are single stranded, positive sense RNA viruses with functionally divided genomes. CMV and BMV are icosahedral viruses based on a $T = 3$ lattice (for definition, see Caspar and Klug, 1962), while AMV particles are bacilliform, probably with $T = 1$ ends (Driedonks *et al.*, 1977). The coat proteins of these viruses are serologically unrelated. They also differ widely in their host and vector specificities. Despite these differences, they exhibit a remarkable similarity of genomic organization. Purified RNA from all the three viruses consist of four RNA segments with approximate molecular weights of 1.3, 1.0, 0.7 and 0.3 million daltons, respectively. RNA 1, RNA2 and RNA4 are monocistronic. RNA4 is the messenger for the coat protein. RNA3 is dicistronic and codes for a non-structural protein; M_r 35,000 called 3a protein of unknown

Abbreviations used: AMV, Alfalfa mosaic virus; CMV, cucumber mosaic virus; BMV, brome mosaic virus; TSV, tobacco streak virus.

function and the coat protein. Apart from these similarities in genomic organization, these viruses have many common physical and chemical properties, which has led Kaper to group these viruses together (Kaper, 1975). Hence, it would be of interest to know if an evolutionary relationship exists between these viruses.

Recently, it has been shown that CMV is evolutionarily related to BMV (Murthy, 1983). In the present study, the amino acid sequence of the non-structural protein of AMV (M_r 35,000; 3a protein) was compared with the sequences of the corresponding proteins of CMV and BMV in an attempt to determine their evolutionary relatedness. The results reveal a significant homology between AMV and CMV or BMV with an evolutionary distance much larger than the corresponding distance between the latter two viruses.

Materials and methods

Complete nucleotide sequence of RNA3 and hence the sequences of the coat protein and the 3a protein of these viruses are now available (Barker *et al.*, 1983; Gould and Symons, 1983; Ahlquist *et al.*, 1981). The 3a proteins of AMV, BMV and CMV are 299,303 and 333 amino acids in length, respectively. In an earlier study, the 3a protein sequences of BMV and CMV were shown to possess a homology of 33.7% (Murthy, 1983). In the present study, the 3a protein sequence of AMV was compared to the corresponding sequences of BMV and CMV by the method of Jukes and Cantor (1969). In this method, segments of length l from the first amino acid sequence are systematically compared to all possible segments of the same length in the second molecule and the homology between them is evaluated. The segment length chosen for comparison was 20 residues. The observed frequency of segments with a high degree of homology was not significantly greater than the expected frequency for unrelated sequences. This could be the case if the sequences are either unrelated or distantly related. However, six segments with good homology (30%) between AMV and BMV or CMV proteins could be identified at AMV residues 1–20, 48–67, 105–124, 157–176, 208–227, 274–293. Accepting these regions as equivalent, the other regions were manually adjusted so as to achieve overlap of residues with high relative substitution frequencies (McLachlan, 1971).

Results and discussion

A total of 264 positions were found to be common for all the 3 sequences. Identical residues at 70 of these positions were found between AMV and either CMV or BMV 3a proteins. This corresponds to a homology of 26.5%. Identical residues were found in all the three sequences at 20 of these positions. The minimum base change per codon (MBC/C) required to convert the AMV 3a protein sequence to the sequence obtained by selecting suitable residues from corresponding positions of BMV or CMV was 0.75. In pairwise comparisons, 281 residues were common between AMV and BMV. 51 of these were found to be identical (18.2%). The MBC/C value for these 281 positions was 1.31. With CMV, 274 common and 39 identical (14.2%) residues were found. The MBC/C value for the 274 positions was 1.31. The final alignments are shown in table 1. 48.0% of the residues are

Table 1. Alignment of the 3a protein sequences of cucumber mosaic, brome mosaic and alfalfa mosaic viruses.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
CMV	-	-	-	-	-	M	A	F	P	G	P	S	R	T	L	T	Q	Q	S	S
BMV	M	S	N	I	V	S	P	F	S	G	S	S	R	T	T	S	D	V	G	K
AMV	M	E	N	T	K	T	N	A	S	S	S	G	M	S	S	S	S	S	F	S
CMV	A	A	L	P	D	D	-	-	-	L	Q	K	I	L	F	S	P	D	A	I
BMV	Q	A	G	G	T	S	D	E	K	L	I	E	S	L	F	S	E	K	A	V
AMV	V	S	Y	A	E	E	M	L	L	A	D	E	V	S	K	I	N	S	M	S
CMV	K	K	M	A	T	E	C	D	L	G	R	H	H	W	M	R	A	D	N	A
BMV	K	E	I	A	A	E	C	K	L	G	C	Y	N	Y	L	K	S	N	E	P
AMV	I	L	G	P	N	Q	L	K	L	C	T	Q	L	V	L	S	N	G	A	A
CMV	I	S	V	R	P	L	V	P	Q	V	T	S	N	N	L	L	S	F	F	K
BMV	R	N	Y	I	D	L	V	P	K	S	H	V	S	A	W	L	S	W	A	T
AMV	P	V	V	L	S	L	V	S	K	E	K	K	S	I	L	N	R	M	L	-
CMV	S	G	Y	D	A	G	E	L	R	S	K	G	Y	M	S	V	P	Q	V	L
BMV	S	K	Y	D	K	G	E	L	P	S	R	G	F	M	N	V	P	R	I	V
AMV	-	-	-	-	-	P	K	I	G	Q	R	M	Y	V	H	H	S	A	I	Y
CMV	C	V	V	T	R	T	G	S	T	D	A	E	G	S	L	K	I	Y	L	A
BMV	C	F	L	V	R	T	T	D	S	A	E	S	G	S	I	T	V	S	L	C
AMV	L	L	Y	M	P	N	I	L	K	S	S	S	G	S	I	T	L	K	L	F
CMV	D	L	G	D	K	E	-	-	-	-	L	S	P	I	D	G	Q	C	V	T
BMV	D	S	G	K	A	A	R	A	G	V	L	E	A	I	D	N	Q	E	A	T
AMV	N	E	A	T	G	E	L	V	D	V	D	T	D	H	D	A	T	Q	A	C
CMV	L	H	N	H	E	L	P	A	L	I	S	F	Q	P	T	Y	D	C	P	M
BMV	I	Q	L	S	A	L	P	A	L	I	A	L	T	P	S	Y	D	C	P	M
AMV	I	F	A	G	R	Y	P	R	S	I	L	A	K	D	A	A	K	G	H	D
CMV	E	-	-	L	V	G	-	-	-	N	R	H	R	C	F	A	V	V	V	E
BMV	E	-	-	V	I	G	G	D	S	G	R	N	R	C	F	G	I	A	T	Q
AMV	L	K	L	V	V	H	A	V	A	S	T	N	A	N	S	A	V	G	V	-

Table 1. *Contd.*

CMV	R	<u>H</u>	<u>G</u>	<u>Y</u>	<u>I</u>	<u>G</u>	Y	G	<u>G</u>	<u>T</u>	T	<u>A</u>	<u>S</u>	<u>V</u>	C	<u>S</u>	N	<u>W</u>	<u>Q</u>	<u>A</u>
BMV	<u>L</u>	<u>S</u>	<u>G</u>	<u>V</u>	<u>V</u>	<u>G</u>	T	<u>T</u>	<u>G</u>	<u>S</u>	V	<u>A</u>	<u>V</u>	T	H	<u>A</u>	Y	<u>W</u>	<u>Q</u>	<u>A</u>
AMV	<u>L</u>	<u>Y</u>	P	<u>I</u>	W	E	D	<u>E</u>	L	<u>S</u>	R	K	<u>Q</u>	<u>I</u>	L	<u>E</u>	R	-	G	<u>A</u>
CMV	<u>Q</u>	<u>F</u>	<u>S</u>	<u>S</u>	<u>K</u>	N	<u>N</u>	<u>N</u>	<u>Y</u>	<u>T</u>	-	-	-	-	-	H	A	A	<u>A</u>	<u>G</u>
BMV	<u>N</u>	<u>F</u>	K	<u>A</u>	<u>K</u>	P	<u>N</u>	<u>N</u>	<u>Y</u>	<u>K</u>	-	-	-	-	-	<u>L</u>	H	G	<u>P</u>	<u>A</u>
AMV	<u>D</u>	<u>F</u>	L	K	F	P	I	A	E	<u>T</u>	E	P	V	R	D	<u>L</u>	L	N	<u>A</u>	<u>G</u>
CMV	<u>K</u>	T	<u>L</u>	<u>V</u>	<u>L</u>	<u>P</u>	<u>Y</u>	<u>N</u>	<u>R</u>	L	A	<u>E</u>	H	S	<u>K</u>	P	<u>S</u>	A	<u>V</u>	A
BMV	<u>T</u>	<u>I</u>	<u>M</u>	<u>V</u>	<u>M</u>	<u>P</u>	<u>F</u>	D	R	L	R	<u>Q</u>	L	D	<u>K</u>	<u>K</u>	<u>S</u>	-	<u>L</u>	K
AMV	<u>K</u>	<u>L</u>	T	D	<u>F</u>	V	<u>L</u>	D	R	<u>T</u>	R	-	-	-	-	-	-	-	-	-
CMV	R	L	<u>L</u>	<u>K</u>	S	Q	L	<u>N</u>	<u>N</u>	V	S	S	<u>S</u>	R	<u>Y</u>	<u>L</u>	<u>L</u>	-	-	<u>P</u>
BMV	N	Y	<u>I</u>	<u>R</u>	<u>G</u>	<u>I</u>	S	<u>N</u>	<u>Q</u>	<u>S</u>	V	D	H	G	<u>Y</u>	<u>L</u>	<u>L</u>	G	R	<u>P</u>
AMV	-	-	-	L	<u>G</u>	<u>V</u>	G	<u>S</u>	<u>K</u>	<u>N</u>	D	P	<u>S</u>	P	V	<u>L</u>	<u>L</u>	E	-	<u>P</u>
CMV	N	V	<u>A</u>	<u>L</u>	<u>N</u>	<u>Q</u>	<u>N</u>	<u>A</u>	<u>S</u>	G	H	E	S	<u>H</u>	<u>I</u>	L	K	<u>E</u>	<u>S</u>	<u>P</u>
BMV	L	Q	<u>S</u>	<u>V</u>	<u>D</u>	<u>Q</u>	V	<u>A</u>	<u>Q</u>	<u>E</u>	D	<u>L</u>	<u>L</u>	<u>V</u>	<u>E</u>	<u>E</u>	<u>S</u>	<u>E</u>	<u>S</u>	<u>P</u>
AMV	R	A	K	<u>I</u>	T	G	<u>K</u>	<u>A</u>	<u>K</u>	<u>T</u>	V	<u>F</u>	<u>I</u>	<u>P</u>	<u>E</u>	G	P	<u>S</u>	<u>V</u>	<u>P</u>
CMV	P	I	<u>A</u>	<u>I</u>	<u>G</u>	<u>S</u>	L	S	A	<u>S</u>	R	<u>N</u>	<u>N</u>	<u>S</u>	F	R	<u>S</u>	<u>Q</u>	<u>V</u>	<u>V</u>
BMV	-	<u>S</u>	<u>A</u>	<u>L</u>	<u>G</u>	<u>R</u>	-	G	V	K	D	<u>S</u>	<u>K</u>	<u>S</u>	<u>V</u>	<u>S</u>	<u>A</u>	<u>S</u>	<u>S</u>	<u>V</u>
AMV	N	<u>T</u>	I	N	<u>G</u>	M	-	-	-	<u>A</u>	P	<u>T</u>	V	<u>R</u>	<u>I</u>	D	<u>A</u>	G	<u>S</u>	<u>P</u>
CMV	<u>N</u>	<u>G</u>	<u>L</u>	C	A	L	L	R	C	<u>T</u>	<u>Y</u>	<u>V</u>	-	<u>Y</u>	<u>I</u>	Y	Y	V	Y	<u>L</u>
BMV	A	<u>G</u>	<u>L</u>	P	<u>V</u>	S	S	P	T	L	R	<u>I</u>	K							
AMV	<u>K</u>	<u>E</u>	<u>L</u>	G	<u>V</u>	P	K	G	F	<u>T</u>	<u>Y</u>	E	S	<u>F</u>	<u>I</u>	K	D	E	I	<u>L</u>
CMV	<u>P</u>	Y	V	N	T	C	E	S	R	V	P	C	E	L	S	T	V	D	I	C
AMV	<u>P</u>	D	H																	
CMV	D	D	A	L	E	D	F	D	H	G	V	S	K	Y	T	S	Q	F		

The residues that are identical in at least two sequences are shown in boxes. Pairs of residues with high relative substitution frequencies are underlines. Single letter codes for amino acids (Schultz and Shirmer, 1979) have been used.

either identical or conservative substitutions (McLachlan, 1971) between AMV and BMV or CMV. The homology observed between these sequences is usually accepted as a definitive indication of an evolutionary relationship.

Van Tol and Vloten-Doting (1981) have observed lack of serological cross-reactivity between the 3a proteins of AMV and the corresponding proteins of CMV, BMV and tobacco streak virus (TSV). However, the same authors note that the 3a proteins might still be functionally equivalent since the coat protein of AMV activates TSV genome, although it does not show serological relationship with TSV coat protein.

In addition to the similarity of genomic organization, AMV, BMV and CMV exhibit similarity in their sensitivity towards sodium dodecyl sulphate, salt and ribonuclease. This has led Kaper (1975) to group them together as viruses stabilized mainly by protein-nucleic acid interactions. The results presented in this paper (table 1) suggest that AMV has indeed diverged from an ancestral precursor of BMV and CMV, and hence Kaper's classification scheme has an evolutionary basis. The x-ray crystallographic investigation of AMV (Fukuyama *et al.*, 1983) and cowpea chlorotic mottle virus (Raymont *et al.*, 1977) which is a bromo virus, are now in progress. A comparison of the structures of the coat protein of these two viruses, when they become available, should provide further understanding of the evolutionary relationship of these two viruses.

Acknowledgements

The authors wish to thank the Council of Scientific and Industrial Research (H.S.S.) and Department of Science and Technology (M.R.N.), for financial support.

References

- Ahlquist, P., Luckow, V. and Kaesberg, P. (1981) *J. Mol. Biol.*, **153**, 23.
- Barker, R. F., Javris, N. P., Thompson, D. V., Loesch-Fries, L. S. and Hall, T. C. (1983) *Nucleic Acids Res.*, **11**, 2881.
- Caspar, D. L. D. and Klug, A. (1962) *Cold Spring Harbor Symp. Quant. Biol.*, **27**, 1.
- Driedonks, R. A., Krijgsman, P. C. J. and Mallema, J. E. (1977) *J. Mol. Biol.*, **113**, 123.
- Fukuyama, K., Abdel-Meguid, S. S., Johnson, J. E. and Rossmann, M. G. (1983) *J. Mol. Biol.*, **167**, 873.
- Gould, A. R. and Symons, R. H. (1983) *Eur. J. Biochem.*, **126**, 217.
- Jaspars, E. M. J. (1974) *Adv. Virus Res.*, **19**, 37.
- Jukes, J. H. and Cantor, C. R. (1969) in *Mammalian Protein Metabolism*, ed. H. N. Munro, (New York: Academic Press), vol. 3, p. 21.
- Kaper, J. M. (1975) in *The Chemical Basis of Virus Structure, Dissociation and Reassembly*, (Amsterdam, New York: North Holland/American Elsevier), p. 1.
- Kaper, J. M. and Waterworth, H. E. (1981) in *Handbook of Plant Virus Infections and Comparative Diagnosis*, ed. E. Kurstak, (Amsterdam: Elsevier/North Holland), p. 257.
- Lane, L. C. (1981) in *Handbook of Plant Virus Infections and Comparative Diagnosis*, ed. E. Kurstak, (Amsterdam: Elsevier/North Holland), p. 333.
- Matthews, R. E. F. (1979) *Intervirology*, **12**, 131.
- McLachlan, A. D. (1971) *J. Mol. Biol.*, **61**, 409.
- Murthy, M. R. N. (1983) *J. Mol. Biol.*, **168**, 469.
- Raymont, I., Argos, P. and Johnson, J. E. (1977) *J. Ultrastruct. Res.*, **61**, 240.
- Shultz, G. E. and Shirmer, R. H. (1979) *Principles of Protein Structure*, (New York: Springer-Verlag).
- Van Tol and Vloten-Doting, L. V. (1981) *Virology*, **109**, 444.